

Diplomski rad br. 658

# Ocjena učinkovitosti operatora križanja u genetskom programiranju

Maja Kontrec

Fakultet elektrotehnike i računarstva

Srpanj, 2014.

## 1 Uvod

## 2 Operatori križanja

- Jednostavno križanje
- Križanje s jednom točkom prekida
- Križanje s očuvanjem konteksta
- Križanje pravedno s obzirom na veličinu
- Uniformno križanje
- Homologno križanje
- Determinističko križanje
- Probabilističko križanje
- Semantičko križanje

## 3 Ispitivanja

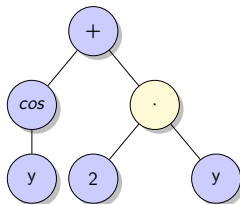
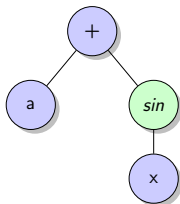
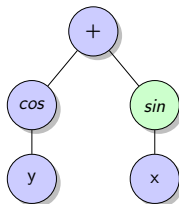
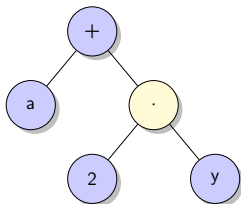
- Problemi za ispitivanje
- Rezultati

## 4 Zaključak

# Uvod - Genetsko programiranje i operatori križanja

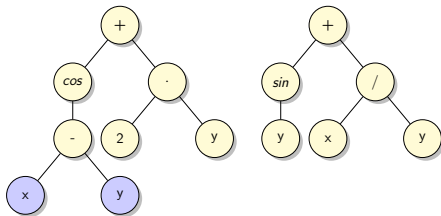
- genetsko programiranje - optimizacijski postupak koji oponaša prirodnu evoluciju
- križanje - ključan element za napredak algoritma
- velik broj različitih operatora križanja
- odabir operatora križanja - čest problem prilikom postavljanja algoritma

# Jednostavno križanje

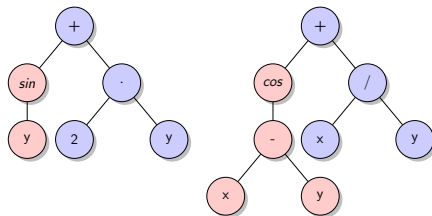


- nasumičan odabir točke prekida u svakom roditelju
- spajanje dva podstabla ispod ili iznad točke križanja

# Križanje s jednom točkom prekida



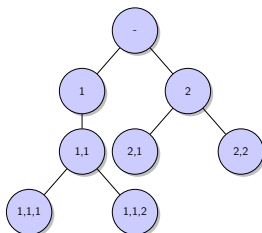
- odabir točke prekida iz zajedničkog područja



- spajanje dva podstabla ispod ili iznad točke prekida

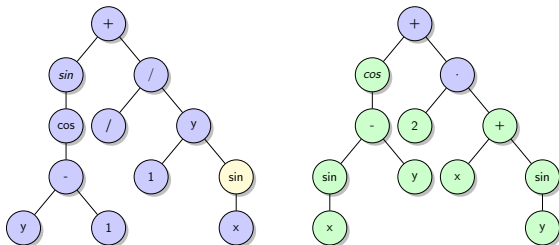
# Križanje s očuvanjem konteksta

- koordinate čvorova stabla



- **križanje s jakim očuvanjem konteksta** - križanje podstabala koja imaju potpuno jednake koordinate korijena
- **križanje sa slabim očuvanjem konteksta**
  - ako su  $T1$  i  $T2$  važeći odabiri podstabala za križanje s jakim očuvanjem konteksta, tada su  $T1$  i  $T2' \subseteq T2$  važeći odabiri podstabala za križanje sa slabim očuvanjem konteksta

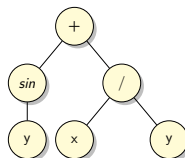
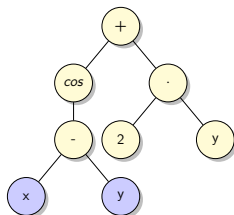
# Križanje pravedno s obzirom na veličinu



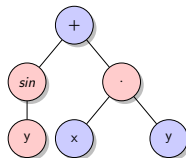
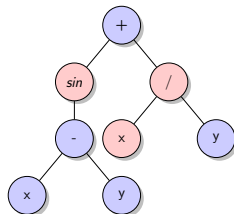
- inačica jednostavnog križanja
- sprječavanje prekomjernog rasta novih jedinki u samom postupku križanja
- odabir nezavršnog čvora kao točke prekida s vjerojatnošću od 0.9
- nasumičan odabir točke prekida  $b$  u prvom roditelju
- u drugom roditelju pronalaze se točke prekida koje su korijen podstabala veličine ne veće od  $1 + 2 \cdot s$ , gdje je  $s$  veličina podstabala u prvom roditelju čiji je korijen  $b$

# Uniformno križanje

## ■ zajedničko područje roditelja



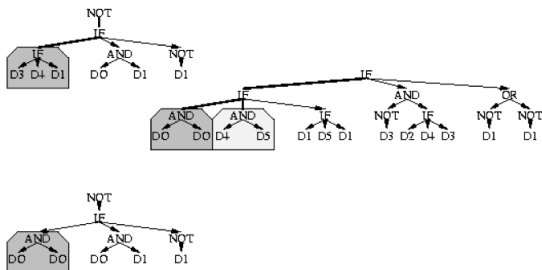
## ■ ravnomjeran odabir čvorova iz dva roditeljska stabla





# Homologno križanje

- uzima u obzir **veličinu i poziciju** podstabala prilikom križanja
- inačica križanja pravednog s obzirom na veličinu



- u drugom roditelju između kandidata koji zadovoljavaju veličinski kriterij prednost imaju podstabla s korijenom slične pozicije kao točka prekida u prvom roditelju

# Determinističko križanje

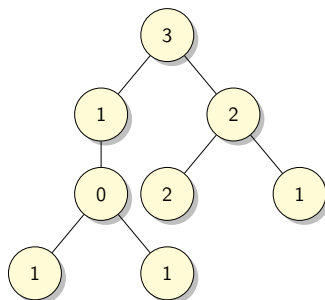
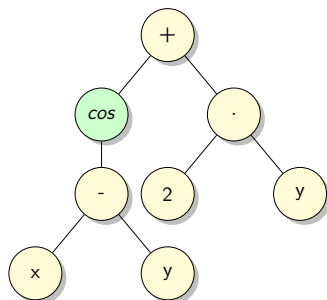
- primjenjivo isključivo na probleme simboličke regresije
- nasumičan odabir točke prekida u prvom roditelju
- u drugom roditelju odabire se točka prekida takva da funkcijska udaljenost između nje i točke prekida u prvom roditelju bude minimalna
- funkcijska udaljenost između dva čvora  $i$  i  $j$  jednaka je

$$d_{i,j} = \frac{1}{2}(|\max_i - \max_j| + |\min_i - \min_j|), \quad (1)$$

gdje su  $\max_u$  i  $\min_u$  maksimalna, odnosno, minimalna vrijednost čvora  $u$  koju on poprima na svim ispitnim primjerima u skupu za učenje

# Determinističko križanje

- primjer:  $x = 1, y = 1$



- primjerice, za čvor označen zelenom bojom, vrijednost  $v$  za dane  $x$  i  $y$  računa se kao:  
$$v = \cos(x - y) = \cos(1 - 1) = \cos(0) = 1$$

# Probabilističko križanje

- probabilistička inačica determinističkog križanja
- primjenjivo isključivo na probleme simboličke regresije
- prilikom odabira točke prekida u drugom roditelju, izračunavaju se vjerojatnosti odabira pojedinog čvora za drugu točku prekida

$$d'_{i,j} = \frac{d_{i,j}}{\sum_{k=1}^s d_{i,k}} \quad (2)$$

$$p_{i,j} = \frac{1 - d'_{i,j}}{\sum_{k=1}^s (1 - d'_{i,k})} \quad (3)$$

- ovime, funkcijski bliži čvorovi drugog roditelja imaju veću vjerojatnost odabira

## ■ logičke funkcije:

$$D = (T1 \cdot TR) + ((1 - TR) \cdot T2) \quad (4)$$

## ■ simbolička regresija:

$$D = (T1 \cdot TR) + ((1 - TR) \cdot T2), \quad (5)$$

gdje je  $TR$  nasumično generirana jedinka,  $T1$  i  $T2$  roditelji, a  $D$  dijete

## ■ programi:

$$D = \text{IF}(\text{CONDR})\text{THEN}(T1)\text{ELSE}(T2), \quad (6)$$

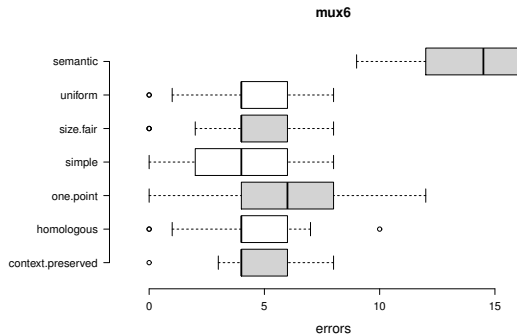
gdje je  $\text{CONDR}$  nasumično generiran program čiji je izlaz interpretiran kao logička vrijednost

# Problemi za ispitivanje

- logičke funkcije
  - 6 i 11 - multipleksor problem
  - evolucija kriptografski sigurnih logičkih funkcija (dviije inačice)
- simbolička regresija
  - klasična simbolička regresija (šest različitih izraza)
  - evolucija funkcije prioriteta za uporabu unutar pravila raspoređivanja
- programi
  - problem umjetnog mrava

- usporedba isključivog korištenja jednog operatora križanja
- u svakom eksperimentu svi parametri osim operatora križanja su konstanti
- zbog eksponencijalnog rasta novih jedinki, semantičko križanje korišteno je u kombinaciji s jednostavnim križanjem u omjeru 1:9

## 6 - multipleksor (minimizacija)

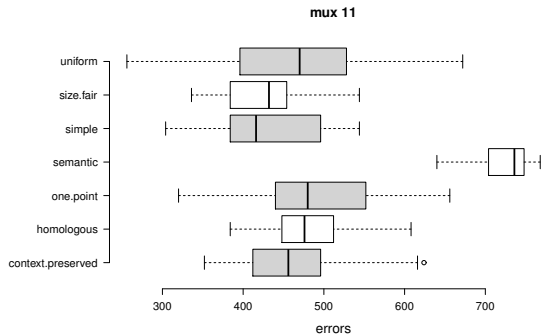


operator	medijan dobrote	rang
simple	4	1
one point	6	6
context preserved	4	1
size fair	4	1
uniform	4	1
homologous	4	1
semantic	14.5	7

- veličina populacije = 400
- broj generacija = 15
- faktor mutacije = 0.4
- nezavršni znakovi: *AND*, *OR*, *NOT*, *IF*
- završni znakovi:  $a_0$ ,  $a_1$ ,  $d_0$ ,  $d_1$ ,  $d_2$ ,  $d_3$



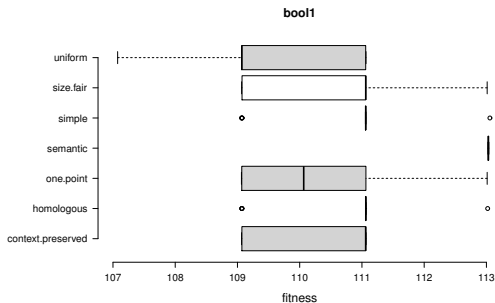
# 11 - multipleksor (minimizacija)



operator	medijan dobrote	rang
simple	416	1
one point	480	6
context preserved	456	3
size fair	432	2
uniform	470	4
homologous	476	5
semantic	736	7

- veličina populacije = 200
- broj evaluacija = 8000
- faktor mutacije = 0.3
- nezavršni znakovi: *AND*, *OR*, *NOT*, *IF*
- završni znakovi:  $a_0, a_1, a_2, d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7$

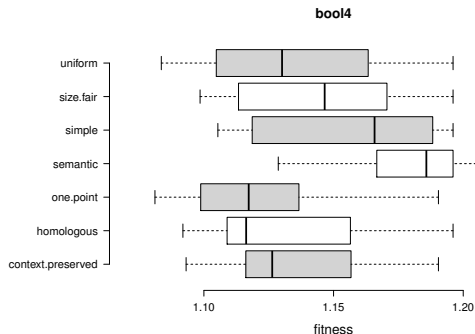
# Evolucija kriptografski sigurnih logičkih funkcija - inačica 1 (maksimizacija)



operator	medijan dobrote	rang
simple	111.061	3
one point	110.0645	6
context preserved	111.06	5
size fair	111.061	3
uniform	109.071	7
homologous	111.062	2
semantic	113.031	1

- dobrota u obzir uzima balans i nelinearnost evoluirane funkcije
- veličina populacije = 250
- broj evaluacija = 3750
- faktor mutacije = 0.3
- nezavršni znakovi: *AND OR NOT IF XOR*
- završni znakovi:  $v_0, v_1, v_2, v_3, v_4, v_5, v_6, v_7$

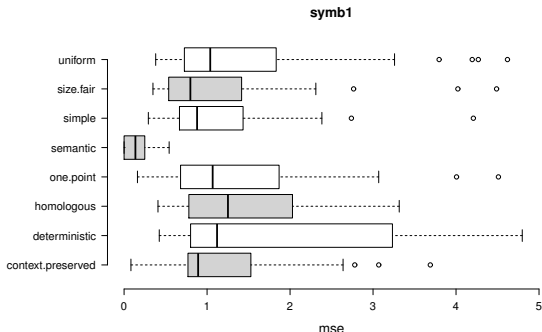
# Evolucija kriptografski sigurnih logičkih funkcija - inačica 4 (maksimizacija)



operator	medijan dobrote	rang
simple	1.16581	2
one point	1.11728	6
context preserved	1.12635	5
size fair	1.14657	3
uniform	1.130145	4
homologous	1.1163	7
semantic	1.185785	1

- dobrota u obzir uzima transparentnost i balansiranost funkcije
- veličina populacije = 250
- broj evaluacija = 3750
- faktor mutacije = 0.3
- nezavršni znakovi: *AND, OR, NOT, IF, XOR*
- završni znakovi:  $v_0, v_1, v_2, v_3, v_4, v_5, v_6, v_7$

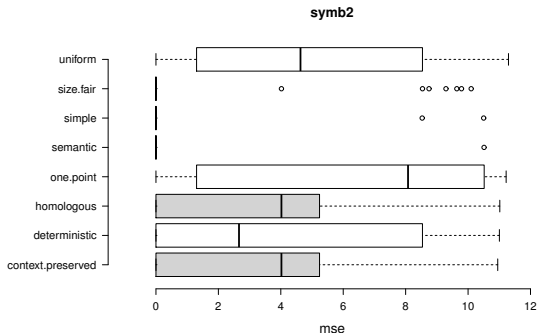
# Simbolička regresija 1 (minimizacija)



operator	medijan mse-a	rang
simple	0.8796885	3
one point	1.06746	6
context preserved	0.8924265	4
size fair	0.7988625	2
uniform	1.038325	5
homologous	1.25174	8
deterministic	1.120405	7
probabilistic	18.16455	9
semantic	0.1364575	1

- izraz:  $\log(x + 1) + \log(x^2 + 1)$ ,  $x \in [0, 20]$
- veličina populacije = 400
- broj generacija = 200
- faktor mutacije = 0.5
- nezavršni znakovi:  $\log$ ,  $+$ ,  $-$ ,  $*$
- završni znakovi:  $x$ ,  $1$

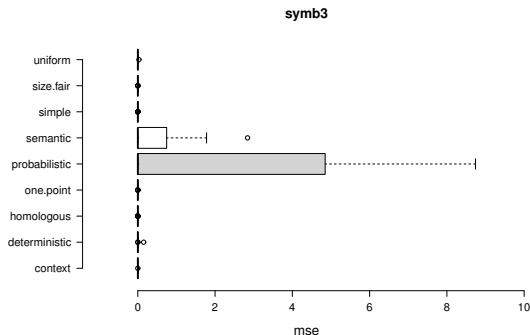
# Simbolička regresija 2 (minimizacija)



operator	medijan mse-a	rang
simple	0	1
one point	8.07678	8
context preserved	4.02037	5
size fair	0	1
uniform	4.6308	7
homologous	4.02037	5
deterministic	2.66101	4
probabilistic	12.3481	9
semantic	0	1

- izraz:  $\sin(x) + \sin(y^2)$ ,  $x, y \in [-10, 10]$
- veličina populacije = 100
- broj generacija = 50
- faktor mutacije = 0.2
- nezavršni znakovi:  $\sin$ ,  $*$ ,  $+$ ,  $-$ ,  $/$
- završni znakovi:  $x$ ,  $y$

# Simbolička regresija 3 (minimizacija)

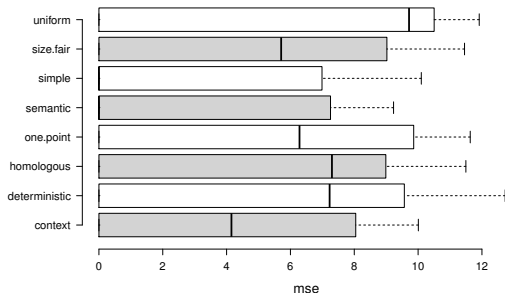


operator	medijan mse-a	rang
simple	$7.77 \cdot 10^{-16}$	1
one point	$7.77 \cdot 10^{-16}$	1
context preserved	$7.77 \cdot 10^{-16}$	1
size fair	$7.77 \cdot 10^{-16}$	1
uniform	$7.77 \cdot 10^{-16}$	1
homologous	$7.77 \cdot 10^{-16}$	1
deterministic	$7.77 \cdot 10^{-16}$	1
probabilistic	$7.77 \cdot 10^{-16}$	1
semantic	$1.39 \cdot 10^{-15}$	9

- izraz:  $2 \cdot \sin(x) \cdot \cos(y)$ ,  $x, y \in [-10, 10]$
- veličina populacije = 400
- broj generacija = 50
- faktor mutacije = 0.3
- nezavršni znakovi:  $\sin$ ,  $\cos$ ,  $*$ ,  $+$ ,  $-$ ,  $/$
- završni znakovi:  $x$ ,  $y$ ,  $2$

# Simbolička regresija 4 (minimizacija)

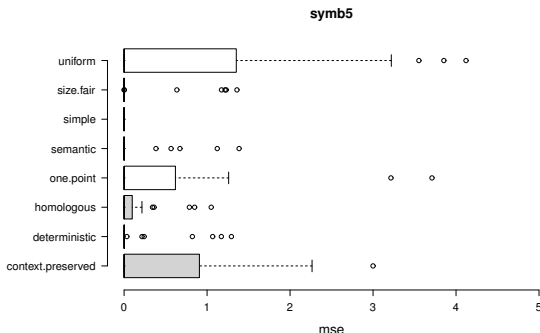
symp4



operator	medijan mse-a	rang
simple	0	1
one point	6.284285	5
context preserved	4.151965	3
size fair	5.709685	4
uniform	9.716085	8
homologous	7.300545	7
deterministic	7.22765	6
probabilistic	12.7275	9
semantic	0	1

- izraz:  $x \cdot y + \sin((x + 1) \cdot (y - 1))$ ,  $x, y \in [-10, 10]$
- veličina populacije = 500
- broj generacija = 200
- faktor mutacije = 0.2
- nezavršni znakovi:  $\sin$ ,  $*$ ,  $+$ ,  $-$ ,  $/$
- završni znakovi:  $x$ ,  $y$ ,  $1$

# Simbolička regresija 5 (minimizacija)

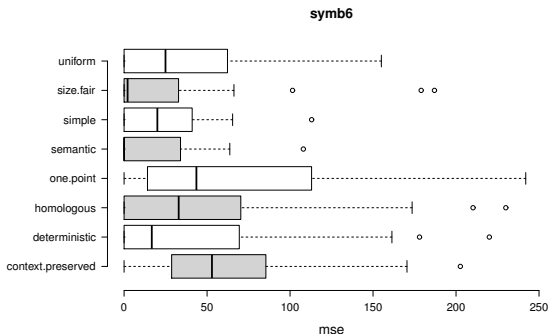


operator	medijan mse-a	rang
simple	0	1
one point	0	1
context preserved	$2.22 \cdot 10^{-16}$	8
size fair	0	1
uniform	0	1
homologous	0	1
deterministic	0	1
probabilistic	7.45469	9
semantic	0	1

- izraz:  $\frac{8}{2+x^2+y^2}$ ,  $x, y \in [-10, 10]$
- veličina populacije = 500
- broj generacija = 200
- faktor mutacije = 0.4
- nezavršni znakovi: +, -, \*, /
- završni znakovi: x, y, 2, 8



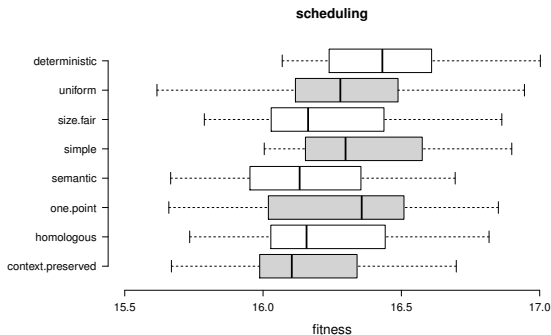
# Simbolička regresija 6 (minimizacija)



operator	medijan mse-a	rang
simple	20.1	4
one point	43.6	7
context preserved	52.9	8
size fair	2.13	2
uniform	25	5
homologous	32.9	6
deterministic	16.7	3
probabilistic	1080	9
semantic	$1.18 \cdot 10^{-13}$	1

- izraz:  $\frac{x^3}{5} + \frac{y^3}{2} - x - y$ ,  $x, y \in [-10, 10]$
- veličina populacije = 300
- broj generacija = 200
- faktor mutacije = 0.4
- nezavršni znakovi: +, -, \*, /
- završni znakovi: x, y, 2, 5

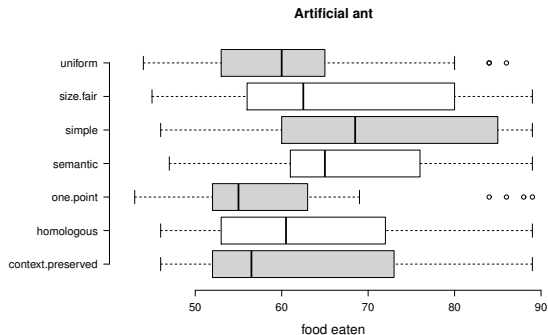
# Evolucija funkcije prioriteta za uporabu unutar pravila raspoređivanja (minimizacija)



operator	medijan dobrote	rang
simple	16.2982	6
one point	16.2566	7
context preserved	16.1039	1
size fair	16.16284	4
uniform	16.2794	5
homologous	16.1527	3
deterministic	16.4313	8
probabilistic	21.30655	9
semantic	16.13225	2

- veličina populacije = 500
- broj generacija = 30
- faktor mutacije = 0.3
- nezavršni znakovi:  $+$ ,  $-$ ,  $*$ ,  $/$ ,  $pos$
- završni znakovi:  $pt$ ,  $dd$ ,  $w$ ,  $SL$ ,  $pmin$ ,  $pavg$ ,  $PAT$ ,  $MR$ ,  $age$

# Problem umjetnoga mrava (maksimizacija)



operator	medijan dobrote	rang
simple	68.5	1
one point	55	7
context preserved	56.5	6
size fair	62.5	3
uniform	60	5
homologous	60.5	4
semantic	65	2

- veličina populacije = 400
- broj generacija = 100
- faktor mutacije = 0.4
- nezavršni znakovi: *Prog2*, *Prog3*, *IfFoodAhead*
- završni znakovi: *left*, *right*, *move*

# Sveukupni rezultati

- rangovi operatora za sve probleme - prosječan rang, ukupan prosječan rang i medijani ranga

operator	prosječan	ukupan prosječan	medijan
simple	2.0833	1	1
one point	5.5	8	6
context preserved	4.1667	4	4.5
size fair	2.25	2	2
uniform	4.4167	7	5
homologous	4.1667	4	4.5
deterministic	4.2857	6	4
probabilistic	7.8571	9	9
semantic	2.8333	3	1

- u većini slučajeva, najučinkovitije jednostavno križanje i njegova kombinacija sa semantičkim u omjeru 9:1
- dosta dobro pokazalo se i križanje pravedno s obzirom na veličinu
- križanje s jednom točkom prekida pokazalo se kao dosta loše, unatoč svojoj širokoj upotrebi
- jednostavno križanje pokazalo se kao vrlo sigurna polazišna točka u odabiru operatora križanja